

AIDS Behav (2010) 14:1124–1136
DOI 10.1007/s10461-009-9635-x

SUBSTANTIVE REVIEW

Adherence and its Measurement in Phase 2/3 Microbicide Trials

Elizabeth E. Tolley · Polly F. Harrison · Els Goetghebeur ·
Kathleen Morrow · Robert Pool · Doug Taylor ·
Stephanie N. Tillman · Ariane van der Straten

Published online: 19 November 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Adherence optimization and measurement have emerged as critically challenging issues for clinical trials of topical microbicides. Although microbicide trials have routinely collected adherence data, their utilization in trial design, implementation, and interpretation has been inconsistent. Drawing on data-driven presentations from several focused meetings, this paper synthesizes lessons from past microbicide trials and provides recommendations for future trials of microbicide and other HIV prevention technologies. First, it describes four purposes for adherence data collection, with particular attention to intention-to-treat versus adherence-adjusted analyses for determining effectiveness. Second, the microbicide field's experiences with adherence measures and data collection modes are discussed, including the strengths and weaknesses of various options and approaches for improving measurement. Then, several approaches to optimizing trial participants' adherence are presented. The paper concludes with a set of recommendations for immediate use or further research.

Electronic supplementary material The online version of this article (doi:[10.1007/s10461-009-9635-x](https://doi.org/10.1007/s10461-009-9635-x)) contains supplementary material, which is available to authorized users.

E. E. Tolley (✉)
Behavioral and Biomedical Research Department, Health and
Development Sciences, Family Health International, 2224 East
Highway 54, Durham, NC 27319, USA
e-mail: btolley@fhi.org

P. F. Harrison · S. N. Tillman
Alliance for Microbicide Development, Silver Spring, MD, USA

E. Goetghebeur
Ghent University, Ghent, Belgium

K. Morrow
The Miriam Hospital and Alpert Medical School of Brown
University, Providence, RI, USA

Keywords Adherence · Microbicides · Clinical trials ·
HIV prevention

Background

In the past decade, six topical, vaginally-inserted microbicide products have been evaluated in one or more clinical trials for their effectiveness in preventing HIV acquisition in women. As yet, none of the trials have produced conclusive evidence of effectiveness. Trials of three products were inconclusive [1, 2, 3], while two products, COL-1492 [4] and cellulose sulfate [5] were found to potentially increase the risk of HIV infection. Most recently, one product [3] has produced a promising, but statistically non-significant result; a recently-completed and larger trial may provide stronger evidence [6].

Given the lack of clear, positive evidence for an effective topical microbicide, adherence has emerged as a critical issue [7]. Yet, improving adherence behavior and its measurement have proven challenging for reasons both unique to the field of topical microbicides, and shared with

R. Pool
Barcelona Centre for International Health Research (CRESIB),
Barcelona, Spain

D. Taylor
Department of Biostatistics, Health and Development Sciences,
Family Health International, 2224 East Highway 54, Durham,
NC 27319, USA

A. van der Straten
RTI International, San Francisco, CA, USA

A. van der Straten
Department of Medicine, University of California, San
Francisco, CA, USA

HIV prevention trials more broadly. First, prevention trials generally recruit healthy, uninfected individuals who may have little incentive to adhere to products of unknown effectiveness [8, 9]. While clinicians ensure product adherence in vaccine or male circumcision prevention trials, product administration and proper dosing in topical microbicide trials is generally participant-dependent and likely to be influenced by participants' understanding, motivations and abilities related to product use. Second, because no validated biomarkers of adherence exist for topical microbicides, adherence measurement has relied largely on self-reports which, for coitally-dependent microbicides, is further complicated by the need to measure associated sexual and condom use behaviors. Finally, the association of these product adherence and risk behaviors with human sexuality and HIV/AIDS—topics that are highly stigmatized in most settings, create additional barriers for both the optimization and accurate measurement of adherence in microbicide trials, as well as HIV treatment and prevention trials more broadly.

In December 2007, the Alliance for Microbicide Development (AMD) and Family Health International convened a 2-day meeting to generate guidance on (1) methods for measuring sexual and product adherence behavior in trial contexts and (2) improving design of clinical trial protocols to optimize adherence to protocol and product use. Drawing upon presentations from the December 2007 meeting, the February 2008 *Microbicides 2008 International Conference*, a November 2006 meeting on *Biomarkers for Evaluating Vaginal Microbicides and Contraceptives* co-sponsored by AMD and CONRAD, and a recent Institute of Medicine report [10], this paper reviews reasons for which microbicide trial adherence data may be collected, highlights issues related to adherence measurement, and discusses approaches to optimizing adherence. It closes with recommendations for adherence measurement and optimization that may have relevance for trials of other female-initiated prevention methods.

Reasons for Collecting Adherence Data in Microbicide Trials

Adherence data are routinely collected in microbicide clinical trials, but their utilization in trial design, implementation, and interpretation has been inconsistent. Because the purpose of measurement will influence the types of data, data collection methods, and/or data sources, it is essential to be clear about the reasons for collecting such data before determining how to measure adherence. Those reasons fall into four categories: (1) determining microbicide effectiveness, (2) providing additional evidence to support trial

results, (3) understanding safety, and (4) understanding acceptability and product experience.

Determining Microbicide Effectiveness

Intention-to-Treat (ITT) Analysis

The central design of Phase 3 microbicide clinical trials assesses the relative reduction in risk of HIV between participants randomized either to an active or a control/placebo product. Primary data analysis typically follows the ITT principle, whereby each participant is assessed according to the treatment regimen assigned at baseline, regardless of post-randomization adherence. If the observed rate of HIV infection is significantly lower in the active than control arm, then one would conclude that the microbicide being tested has the intended biological effect of reducing risk of infection, as long as it could be assumed that the rate of HIV exposure (which is influenced by risk-taking behaviors and alternative protective measures) had been comparable between study arms (i.e., as would be hoped for in a randomized study with a blinded control arm).

ITT analysis is considered the gold standard for analyzing randomized clinical trial data because it provides the basis for an unbiased test of an effect of treatment regimen. However, because ITT analysis ignores risk-taking behavior, many consider the generalizability of its conclusions to a broader population under routine conditions to be deficient. Consider the following: (1) If the trial organizers manage to achieve 100% perfect condom use, there should be no vaginal transmission of HIV and the investigators will be unable to demonstrate any effect of the microbicide—positive or negative. (2) If product use and/or risk behavior differ between arms, such as the differential condom use seen in the “Methods for Improving Reproductive Health in Africa” (MIRA) trial [11], then the ability to make robust statements about the microbicide's biological effect will be undermined. (3) If adherence happens to be particularly poor during acts where HIV exposure takes place, then even if the test product is highly efficacious when used as intended, the reduced ITT effect will not be detected, even if adherence is quite high overall. Hence, it is imperative that adherence and risk-taking behaviors be observed or reported and documented accurately, and their implications for the valid interpretation of the ITT analysis examined.

Pre-Planned Adherence-Adjusted Analysis

Ideally, if it were possible to know each participant's infection status before and after each sex act with an infected partner, and accurately record the presence or absence of condom and/or microbicide use for each act,

then an adherence-adjusted analysis would provide a more powerful approach than an ITT analysis. However, because the infection status of sexual partners is generally unknown and participants' infection status is assessed only infrequently, such detailed adherence-based analysis is essentially impossible. Investigators must depend instead on measures of adherence and risk-taking behavior that can be applied in those periods between HIV test visits, periods of varying extent during which the frequency and type of exposures to HIV cannot be readily known. Nevertheless, in theory, if a trial could reliably identify women likely to be adherent prior to their randomization (e.g., based on data collected during a "run-in" period or on some other validated predictor of adherence), adherence-adjusted analysis could still provide convincing, randomization-based evidence of a test product's effectiveness in a subgroup, even if the ITT analysis failed to demonstrate a protective effect overall [12]. (An appendix, available online, describes the potential for a two-stage biomarker-adaptive design for microbicide trials.) The potential benefits of such a pre-planned adherence-adjusted analysis justify efforts to identify valid predictors and measures of adherence.

Providing Additional Evidence to Support Trial Results

Despite continuing debate about whether adherence data should be included in primary analyses of effectiveness, most agree that such data can provide valuable information to support or explain trial results. For example, if ITT analysis were to conclude that a new microbicide product was effective, an unconfounded per-protocol, "on-product" analysis would likely find a still higher level of effectiveness, providing insight into the product's potential efficacy when used consistently and correctly. Similarly, if ITT analysis found a suggestive although non-significant reduction in risk—but a noticeably larger effect was observed in an adherent subset (as occurred in the HPTN 035 trial), then the investigators might be motivated to conduct a larger, more powerful trial.

In contrast, should an ITT analysis conclude no difference in infection rates between the active and placebo arms, several substantially different explanations may exist. If reported risk-taking behavior were comparable across arms, and if adherence were found to be high, then the results were more likely due to a lack of meaningful product efficacy. On the other hand, if risk factors were not comparable between arms (as appears to have been the case with condom use in the MIRA diaphragm trial) [11], then it may be difficult or impossible to assess product efficacy. Likewise, if otherwise highly-adherent women tended to avoid product use in situations when the virus were most likely to strike (e.g., when having sex with an infected

primary partner), then the product's potential action was never tested and no information on efficacy obtained. In the latter case, even perfect measures of adherence might not provide sufficient information for drawing unequivocal conclusions on the relationship between product adherence and efficacy.

In sum, pre-planned adherence-adjusted analyses that are not randomization-based generally carry a risk of misleading conclusions. This is acutely the case for clinical trials of microbicides, where product use is associated with varying and unpredictable patterns of sexual and risk-taking behavior. This does not mean that adherence-adjusted methods should never be undertaken—just that they must be used with caution and, where possible, predicated on structural or other randomization-based approaches to analysis [13].

Understanding Safety

Evaluating safety is a primary aim of Phase 2 microbicide trials and a secondary aim of Phase 3 trials. Longer-term use of any investigational product may affect human safety as either a direct result of harmful side effects or, indirectly, as a result of changes in participant behavior. For instance, secondary analysis of adherence data from the trial of COL-1492, an N-9-based gel, indicated that the product increased the risk of HIV infection in women who used it more than 3.5 times daily and who had a high incidence of lesions with epithelial disruption [13]. Product use in the context of effectiveness trials may also contribute to negative safety results indirectly if participants, believing that the test product protects them from HIV infection, lower their pre-trial levels of condom use or modify other risk-reduction behaviors [10].

Monitoring Acceptability and Optimizing Adherence

Examining patterns of adherence in the context of an ongoing clinical trial can lead to insights into product acceptability that are valuable for the trial itself, for the design of future trials, and for eventual product introduction and utilization. Scrutiny of adherence patterns can reveal who can and cannot use microbicides correctly and consistently, and what factors enhance or constrain such use. Better understanding of trial participant perspectives on the acceptability of the test product and their experience with it can help implementers optimize adherence as a trial proceeds. Such understandings will also be relevant to future product introduction and delivery, and may contribute to fashioning individualized prevention strategies and identifying and responding to key marketing niches. Finally, acceptability and adherence data from ongoing and completed trials can contribute importantly to more realistic designs for future trials.

Measuring Adherence

A number of adherence measures have been or could be used in microbicide clinical trials. These fall into three general categories. *Direct* measures of adherence, often referred to as “biomarkers”, are substances or effects whose presence or absence indicates that a biological or pharmacological process has occurred in response to a drug. *Indirect* measures of adherence comprise two major sub-categories: “objective measures” and “self-report measures”, both reliant on the observations or reports of clinicians, trial participants, or others. Table 1 presents an overview of these measures, by category and data collection mode, with their defining characteristics and the purpose(s) to which they best apply. Below, we describe these three categories of adherence measures in greater detail, followed by a brief discussion of how data collection modes may influence self-reported adherence measures, and concluding with potential strategies for improving adherence measurement.

Direct Measures of Adherence

Development of direct, respondent-independent, quantitative biological measures—“biomarkers”—of adherence and incident infection might enable the kind of unbiased, adherence-adjusted effectiveness analyses described in “[Reasons for Collecting Adherence Data in Microbicide Trials](#)”. Unfortunately, progress in developing, validating, and harmonizing such tools has been limited, despite the urgent need for them in microbicide and other areas of basic, translational, and clinical research [14, 15, 16].

Biomarkers of Semen Exposure

Prostate-specific antigen (PSA), semenogelin (Sg), and Y-chromosome DNA (Yc DNA) have been incorporated into microbicide-related sub-studies as possible tools for validating self-reported data on condom use and sexual activity. Ideally, such biomarkers would be consistently detectable in the female reproductive tract when exposure to semen had occurred; have low variability in concentration levels, physical distribution and time to clearance; should not be affected by the microbicide under study or other factors present in the female reproductive tract; and should be stable, sensitive, specific, and feasible in a variety of settings. Their potential for measuring adherence is presently limited by the fact that semenogelin is detectable in vaginal fluid samples for only up to 3 days, PSA only up to 48 h. Furthermore, non-detection of semen does not necessarily mean that a condom was used during intercourse; it could also mean that intercourse had either occurred outside the period of detection or had not occurred at all [17, 18].

Applicator Tests

In its Phase 3 trial of Carraguard®, the Population Council based its primary measure of product adherence on a dye test intended to indicate whether an applicator had come into contact with mucins, mucoproteins that are characteristically expressed by female reproductive tract epithelial cells. Participants were asked to return all applicators distributed to them, whether or not they had been used; these were then laboratory-processed to reveal a blue stain if mucin contact had occurred [17]. While such tests could at least theoretically provide a more objective indicator of whether returned applicators had been used, they cannot provide information on the timing of use, the amount of product inserted, or which mucosal surfaces had been touched by the applicator. Furthermore, the reliability of dye tests may vary, depending on the material composition or form of the applicator.

Drug Level Assays

As work advances on anti-retroviral (ARV)-based topical microbicides, more adherence measurement options may become available. Some ARV-based gels are detectable for days or even weeks after insertion but, as with the Population Council’s applicator test, they cannot indicate the timing of gel insertion in relation to sex, or whether other coital acts occurred without gel use [17].

Additional Biologic Measures

The International Partnership for Microbicides (IPM) is developing a bar-coded “Smart applicator” that would register time and date of use and the ambient vaginal temperature when the applicator was inserted, and collect mucin to stain for presence of vaginal secretions but again, could provide no record of the number of coital acts in which no applicator—and therefore no product—was used. IPM is also exploring the potential of a “Sexometer” that would measure microbicide use during coitus; future prototypes could, theoretically, be designed with sensors to measure the presence of gel, semen, and virus to determine exposure to HIV [16]. In sum, despite efforts to develop and validate biomarkers for use in microbicide clinical research, technological and practical challenges continue to limit their use.

Indirect, Objective Measures of Adherence

Measures are considered “objective” when they are independent of individual feelings, beliefs, or desires and do not involve self-report from the subject/participant. In the case of regimens for product use in clinical trials, such

Table 1 Adherence measures considerations: purposes within clinical trials, strengths, and biases

Measure	Purpose	Best match to product type				Sources of measurement strength							Good sensitivity	Good specificity	Analysis of patterns
		Effectiveness outcomes	Explanatory of CT results	Safety information	Acceptability information	Added to existing assessment	Respondent independent	Low tech or low cost	Ease of admin or low Subject burden	Subject to individual differences in absorption, metabolism, etc	Inadequate sensitivity	Inadequate specificity			
Direct, biomarker measures															
HIV test (antibody or viral particle)	✓			✓		✓	✓		✓	✓		✓			
Drug levels (plasma or genital tract)	✓		✓	✓		✓	✓			✓		✓		✓	
Semen biomarkers		✓	✓	✓		✓	✓			✓ ^e		✓		✓	
Applicator dye test		✓	✓	✓		✓	✓	✓ ^e		✓ ^e		✓ ^e		✓ ^e	
Indirect, objective measures															
DOT (directly observed therapy)	✓		✓	✓	✓		✓							✓	
MEMS cap—product count			✓	✓	✓ ^e		✓ ^e		✓			✓ ^e		✓	✓ ^e
Pharmacy disbursements		✓	✓	✓	✓	✓	✓		✓			✓			
Condom returns		✓	✓	✓	✓		✓		✓ ^e			✓ ^e			
manual product count		✓	✓	✓	✓ ^e		✓								
Indirect, self-report measures															
Self-report (face-to-face interview; structured or quantitative)		✓		✓	✓			✓				✓ ⁱ		✓	
Self-report ([AJCASI)		✓	✓	✓	✓	✓			✓			✓ ^e		✓	
Self-report (Diary)		✓	✓	✓	✓	✓		✓		✓ ^{fl}		✓ ^e		✓	
Self-report (face-to-face interview; in-depth qualitative)		✓	✓	✓	✓	✓				✓		✓ ^{ech}		✓ ⁱ	
Sources of measurement bias or limitation															
Indirect measure	Addition to existing assessment	Respondent dependent	High tech or high cost	Difficult to admin or high subject burden	“Negative” result has many interpretations	Under- or over estimates adherence or does not capture relevant exposure periods	Chain of custody concerns	Subject to response demand characteristics	Only applicable to drugs that are absorbed	Subject to individual differences in absorption, metabolism, etc	Inadequate sensitivity	Inadequate specificity			
Direct, biomarker measures															
HIV test (antibody or viral particle)			✓		✓	✓									
Drug levels (plasma or genital tract)			✓	✓		✓			✓						
Semen biomarkers	✓		✓	✓	✓	✓		✓							
Applicator dye test	✓				✓	✓	✓								
Indirect, objective measures															
DOT (directly observed therapy)	✓		✓	✓		✓									
Mens cap – product count	✓	✓	✓		✓	✓							✓		
Pharmacy disbursements		✓			✓	✓	✓	✓					✓ ^e		
Condom returns	✓	✓	✓	✓	✓	✓	✓	✓					✓		

Table 1 continued

Measure	Sources of measurement bias or limitation												
	Indirect measure	Addition to existing assessment	Respondent dependent	High tech or high cost	Difficult to admin or high subject burden	“Negative” result has many interpretations	Under- or over estimates adherence or does not capture relevant exposure periods	Chain of custody concerns	Subject to response demand characteristics	Only applicable to drugs that are absorbed	Subject to individual differences in absorption, metabolism, etc	Inadequate sensitivity	Inadequate specificity
Manual product count		✓ ^e		✓	✓	✓	✓						
<i>Indirect, self-report measures</i>													
Self-report (face-to-face interview; structured or quantitative)	✓	✓	✓			✓ ^e	✓		✓			✓ ^e	
Self-report (IA/CASI)	✓	✓	✓	✓ ⁿ	✓ ^e		✓ ^e		✓				
Self-report (Diary)	✓	✓	✓ ^m		✓ ^e	✓	✓		✓			✓ ^e	
Self-report (face-to-face interview; in-depth qualitative)	✓	✓	✓	✓ ^j ; ^k	✓				✓ ^e				

^a Any formulations requiring an applicator to insert into a cavity with mucosal surfaces, e.g., topical gels^b Oral tablets^c Intravaginal Ring (IVR) requires clinician confirmation^d IVR if packaging can be equipped for MEMS^e Level of characteristic depends on study protocol: e.g., whether participants are required to bring items back to clinic or staff go to participants' homes; whether data are collected continuously or cross-sectionally; whether data are collected on the entire sample or a subsample and how that subsample is determined; whether data are triangulated with other measures to achieve greater accuracy through reconciliation of inconsistencies^f Best if done as unscheduled visit^g Highest costs for (A)-CASI; PDA; qualitative, depending on (sub)sample size^h Especially if triangulatedⁱ Able to capture subtle nuances of product use; can tailor administration to assist recall^j Training time^k Analysis time^l Privacy; or Audio minimizes literacy requirement^m Some degree of literacy requiredⁿ Electricity required; technology safety required

measures have included applicator and pill counts and electronic drug-monitoring systems.

Applicator counts have been defined as “objective” because they do not rely on participant self-report. Yet such counts are not totally free of bias since they rely on clinicians’ accuracy in counting, calculating, and recording product-use data, and their interpretations of whether or not an applicator has actually been used. Applicator counts also depend on participants’ willingness to transport products between study clinic and home, and on availability of storage in often spatially constrained domestic environments with limited privacy. Applicator counts can also generate under- and over-counting of product use; for example, counting returned empty gel applicators as proxies for use-adherence may *over-count* adherence, while interpreting failure to return unused applicators as proxies for adherence may either *over-* or *under-*estimate it.

Pill counts have similar limitations. Studies of adherence to antiretroviral therapy have concluded that pill counts may *overestimate* adherence if participants either forget or deliberately fail to bring back all unused product; electronic drug-monitoring systems like the Medication Event Monitoring System (MEMS) may *underestimate* adherence if participants remove more than one pill from the bottle at a time. It is also possible that if the MEMS cap is lost or damaged, data could be compromised in an unpredictable direction [19, 20].

Indirect, Self-Reported Measures of Adherence

Adherence measurement has more typically been based on self-reported data and assessed post-randomization. Self-reported data can be collected through a range of techniques, including face-to-face interviews (FTFI), self-administered paper or computer-based questionnaires, diaries, or telephone and text-messaging options. Although all self-reported data are influenced by respondents’ ability to recall and their willingness to provide accurate information, choices about data collection mode and question development may bias responses in different ways. In addition, adherence questions may vary in terms of time reference, response format, and/or level of structure.

Time Reference

An individual’s ability to compute estimates of sexual behavior is influenced by the frequency of a given behavior, the reporting time frame, and the vividness and complexity of the behavior itself [21]. To date, most microbicide trials have measured product adherence at last sex, or as the proportion of sex acts covered during the past week, rather than the entire 1- to 3-month period between scheduled follow-up visits, assuming that the more

proximal the time reference, the more accurate participants’ recall of events.

While a standard approach to adherence measurement can facilitate comparisons across trials, such truncated time references may inevitably not be optimal. For example, the extraordinarily high levels of gel use (97% use at last sex) reported in the Carraguard[®] trial were not confirmed by the applicator tests, suggesting that participants may have given “socially desirable” responses or altered behaviors prior to a forthcoming clinic visit (the “white coat” effect) in order to align behaviors with reporting expectations [22]. A review of baseline data from seven trials and 27 sites found wide variation in the mean number of reported sex acts per week, from one sex act per week in rural South Africa to almost 27 sex acts per week in Kampala, Uganda [23]. In populations with high reported coital frequency, a 1-week time frame may be too long for accurate recall. Conversely, where coital frequency is low, a 1-week time frame may be too short to accumulate the number of acts needed for calculating the proportion of risky sex acts.

Exit interviews can reduce response burdens on participants and attenuate compulsion toward socially desirable responses. As one example, 9–15% of MIRA participants reported at exit that they had ever over- or under-reported diaphragm and/or gel use in the course of that trial. Still, while exit interviews do help validate data obtained through other self-report methods, collecting information solely at exit presents participants with an increased level of difficulty: having to accurately recall both specific behaviors (i.e., sex and product use) and the factors that had influenced them [24].

Response Format

Self-reported adherence questions can be worded as frequencies (e.g., number of gel applications in the last week), proportions (gel use as a proportion of sex acts), binary categories (use/non-use), or Likert-type scales (e.g., always, sometimes, never). While most microbicide trials have tied adherence measures to specific sexual episodes, one sub-study complemented the proportion-based measure of condom and gel adherence with a second measure based on participants’ *estimates* of consistency over a longer (2-month) time period. Participants were asked whether they had ever missed using a condom/study gel during sex and, if so, how often, according to a five-point scale (never, rarely, sometimes, frequently, always). While respondents may liberally interpret Likert-type categories [23], this sub-study found that the scale of perceived consistency better differentiated between levels of condom adherence than did the single measure of condom use as a proportion of sexual acts. At the study’s 2-month follow-up visit, based on the measure of condom use as a proportion of sex acts in

the past week, 72% of participants reported 100% condom use. In contrast, based on the scalar approach, just 42% of those same participants reported always using a condom, 29% reported sometimes or frequently using a condom, and 29% reporting rarely or never using a condom over the longer 2-month time frame [25].

Degree of Structure

In-depth interviews (IDIs), focus group discussions (FGDs), and other qualitative methods are well suited to investigating complex behaviors and what influences them. Indeed, FGDs conducted with participants and their partners exiting from the MIRA trial helped validate quantitative findings and generated information that proved critical to understanding differential condom use between study arms [26]. Nevertheless, qualitative methods generally require special training and tend to be time-consuming for participants, researchers, and those who must analyze the resulting data. Furthermore, they do not address clinical scientists' preferences for standardized instruments that support comparisons across trials, time periods, and subpopulations.

Psychosocial scales comprising multiple questions or statements measuring an underlying or latent theoretical construct (e.g., a measure of protective efficacy or adherence intent) could provide the standardization and ease of administration desired by clinical researchers; however, they require adequate pre-trial studies to linguistically and culturally translate existing measures or develop new ones. Rigorous methods have been developed for this task, including: (1) qualitative interviews to generate items and relevant response sets; (2) cognitive interviews to assess the ordering, framing, and wording of items; and (3) psychometric evaluation to determine scaling and validity. Once developed, these measures should be assessed for use in multiple settings, to be validated across populations and so that cross-trial data can be accumulated for future meta-analyses.

Data Collection Modes for Self-Reported Measures

Choice of data collection mode may improve recall, reduce social desirability bias, or both. Coital and/or product adherence diaries provide one relatively low-tech approach to measuring adherence. Computer and telephone-assisted interview techniques are other potential approaches. All have their pros and cons.

Diaries can provide a more complete, continuous record of how products are used during trials. Yet diary-based data collection is renown for substantial limitations: the need for some level of participant literacy or numeracy; high participant burden; and a documented tendency by participants

to record information in a lagged rather than daily manner, often just before a study visit.

Audio Computer-Assisted Self-Interview (ACASI) techniques increase privacy and are therefore expected to decrease social desirability bias, thereby enhancing data quality and reliability. An interview mode experiment ancillary to the Carraguard[®] trial sought to test this assertion. The sub-study compared participants' reports of sexual and adherence behavior, who had been randomized either to ACASI or FTFI. Preliminary analysis found significantly higher levels of reported sexual risk behaviors (e.g., number of sexual partners and anal sex) by ACASI compared to FTFI. In contrast, there were no differences by data collection mode in the proportion of women reporting condom or gel use at last sex, reports of condom use in the prior 2 days, or positive tests for the presence of semen [27]. Given the required and therefore routinized use of condoms and gel in a microbicide trial, it is possible that data collection mode had little bearing on how participants reported their adherence behavior, whereas the privacy afforded by ACASI enhanced reporting of more stigmatized sexual behaviors.

Despite the potential for increased, possibly more accurate reporting of some behaviors, ACASI is not without difficulties. Separately, some MIRA trial participants reported confusion about questions and admitted wanting to complete ACASI questions quickly, suggesting that participant comprehension and attention may be challenged by this method and require someone who can help the participant with clarifications and maintenance of focus [24]. A pictorial-based application of ACASI, planned for the "Vaginal and Oral Interventions to Control the Epidemic" or VOICE trial [28], may avoid such problems. In this study, which aims to evaluate the safety and effectiveness of tenofovir (Viread[®]) and Truvada[®] taken orally as pre-exposure prophylaxis (PrEP) and tenofovir gel applied vaginally, participants will respond to a sub-set of questions about sexual behavior and adherence that appear on the screen in local language text and images, enabling low-literacy women to respond to key inquiries about number of sexual partners and tablets, gel, and condom use.

Interactive Voice Response (IVR) technology is another innovation employed in a Phase 1 microbicide safety study to collect daily adherence diary data [29]. IVR permits callers to dial a telephone number that is answered by the IVR system. A pre-recorded or dynamically-generated voice explains the interview options and administers interview questions to which the participant caller responds by pressing numbers on the telephone keypad or speaking them into the telephone. Callers are reminded about their compensation, which is accrued and automatically tallied at the end of each call. The convenience and privacy possible through IVR may increase the feasibility of

obtaining daily adherence data at the same time that it decreases social desirability bias. However, gains in privacy, more frequent follow-up, and quicker access to data must be balanced with the high cost of programming, potential loss of data, and lack of participant familiarity and comfort with technologically complex data collection methods.

Strategies to Improve Adherence Measurement

Various attempts have been made to enhance and assess the accuracy of adherence data. These include “mixed-method” approaches, including triangulation procedures; development and application of composite measures; and identification of baseline adherence predictors.

Triangulation refers to the use of multiple observers, theories, or data collection methods to overcome the inherent biases of any single observer, theory, or method, increase convergence and reconcile inconsistencies across data sets. Social scientists used triangulation to identify adherence-related patterns and problems in the Microbicides Development Programme (MDP) trial 301 in sub-Saharan Africa. At each of three visits, adherence data from case record form (CRF) interviews, applicator returns, and coital diaries were collected on a random subset of 100 women per site and entered into a comparison form. A few days later, each woman participated in an IDI during which any inconsistencies in the comparison form were probed. While some inconsistencies were identified in over half of forms, women provided plausible explanations for the majority of discrepancies, indicating that inaccurate reporting was usually unintentional. All these data, with data from partner interviews, ethnographic research, and focus group discussions, were entered into a summary database, coded, analyzed, and reported back to the trial managers [30].

Composite measures are combinations of different measures to generate a single outcome whose value falls somewhere between outcomes from individual measures. Composite measures may be generated from a single data collection method (e.g., responses to a series of self-reported adherence questions) or from different data collection methods (e.g., a biomarker, product count, and a self-reported measure.) The ADEPT (Adherence and Efficacy to Protease Inhibitor Therapy) study, a prospective observational investigation of adherence to medication for HIV suppression, examined the utility of such a measure. The study sought to determine how adherent participants would be when initiating HAART, the relationship between adherence and virologic outcomes, any psychological factors that might predict adherence, and how different adherence measures compare with each other and, possibly, predict virologic outcomes. A composite

adherence score (CAS) was developed based primarily on data from electronic medication (MEMS®) bottle caps containing a microchip recording each instance of bottle opening; missing data were supplemented with pill counts and self-reported adherence when MEMS data were missing or inaccurate. The CAS indicated levels of adherence higher than MEMS and lower than self-reported or pill count and, most importantly, had a stronger correlation with the primary objective measure, viral load suppression, than any of its single components [20].

Baseline adherence predictors, as described earlier, could provide valuable information to support or explain trial results. Measuring an adherence predictor at baseline helps preserve the role of randomization for an adjusted analysis, thus ensuring that non-product-related predictors of HIV infection are distributed in a balanced way across study arms. Once measured, baseline adherence predictors could be applied in secondary/confirmatory analyses to identify and remove participants who were likely to have been non-adherent; removing even 5% of non-adherent participants could greatly improve statistical power to detect a treatment effect [31].

Examples of measures that might qualify as baseline predictors include: an observed level of adherence to a comparable product collected during a study run-in, a validated self-efficacy scale measured at screening, or the recorded number of product insertions attempted by a participant at her enrollment visit. Unfortunately, little research has been conducted to identify and evaluate potential baseline indicators of product adherence in clinical trials. The design of the forthcoming Fem-PreP trial, which will examine the safety and effectiveness of a once-daily Truvada® pill taken prophylactically to prevent HIV transmission in women, includes a 2- to 4-week-long vitamin run-in phase prior to randomization [32]. One purpose of the run-in is for participants to practice pill-taking and to allow for tailored counseling on adherence at enrollment based on the participant’s experience. Another purpose is to measure the ability of potential participants to swallow a vitamin pill similar in size to Truvada® in front of staff at enrollment. Women who cannot do so will be excluded. An evaluation of this baseline intervention could provide valuable information about whether such run-in behavior would or would not predict actual trial behavior.

Optimizing Adherence

There are several points at which adherence can be optimized: (1) a priori, in both the overall design of the trial and as explicitly crafted elements of that design; (2) in trial implementation, beginning with recruitment, screening, and enrollment; and (3) during the trial, in response to

monitoring indicating that adjustments are required. Together, these somewhat overlapping categories indicate that optimization of adherence must be both a very early consideration and an ongoing dimension of trial implementation.

Optimizing Adherence in Trial Design

CAPRISA 004

To date, at least one ongoing microbicide trial—the CAPRISA 004 Phase 2B trial of tenofovir—has developed explicit strategies for optimizing adherence prior to study initiation. The trial's Adherence Support Program (ASP) employs “job aids” (flip charts, information leaflets, and clock/calendar materials) and carefully constructed messages to support the provision of personalized adherence counseling to trial participants [33]. CAPRISA participants are also given diaries in which to record trial-related and other information. Though optional, many women reportedly consult their diaries when responding to behavioral questions during clinic visits. A formal evaluation of the ASP is not currently envisioned, but routine feedback from the clinical trial team suggests that its personalized approach to counseling is time well spent.

Optimizing Adherence in Trial Implementation

Adherence Capacity

One approach to optimizing adherence would be to identify and recruit the “right” people into the trial, that is, those who appear to have both the appropriate risk profile as well as “adherence capacity”. The run-in scenario used for the Fem-PreP trial design, described in “[Strategies to Improve Adherence Measurement](#)” above, offers one strategy for identifying participants who may be more likely to adhere to daily pill use through (1) observation of their pill-taking abilities during enrollment and (2) assessing participants' level of adherence during a run-in period. Still, without prior evidence that this indicator of adherence capacity could constitute a valid predictor of product use and adherence in a trial context, clinical researchers may be reluctant to predicate participant selection solely on such a measure.

Optimizing Adherence in Response to Routine Monitoring

HPTN 035 As the HPTN 035 effectiveness trial of BufferGel and PRO 2000 gels got under way, monitoring efforts flagged lower than targeted levels of gel adherence. Thus, in early 2006, the protocol team met with

clinic staff to explore and address reasons for gel non-use. These conversations revealed that some clinic staff were conflating training messages that stressed the possibility of gel side effects, the fact that gel efficacy against HIV infection was unknown, and the proven effectiveness of condoms against pregnancy and HIV infection. As a result, participants were receiving the following advice: *Since the gels may have side effects and may not protect against HIV, only use the gels when you are also able to use condoms.* After consultation, the group revised its adherence messages and developed associated scripts for all sites to accord with the following message: *In order to properly test if the gel protects against HIV, it is important that you use your gel during every sex act, even when condom use is not possible.* As a result, gel use, overall and as a percentage of sex acts in which condoms were also used, steadily increased, with the biggest gain in gel adherence for sex acts in which a condom could not be used [34]. This experience emphasizes the importance of harmonizing risk reduction, contraception, and product counseling, and ensuring that clinical trial staff understand the associated messages thoroughly in order to convey them accurately and persuasively to trial participants.

Trajectory analyses, a statistical approach derived from developmental psychology that reveals individual or group behavioral patterns over time, may provide a useful approach to monitoring adherence and identifying those requiring assistance [35]. For example, while trajectory analyses of self-reported gel adherence data from Savvy Ghana and CONRAD CS trials identified a substantial cluster of women reporting high, sustained gel use, other clusters exhibited suboptimum adherence, including approximately 10% of women reporting initially low but increasing adherence, and 10–20% reporting declining gel use [36]. And, when the high sustained condom users identified in analysis of the Savvy Ghana trial were divided into those reporting “perfect” (reported condom use always equal to the number of sex acts in a given time period) and “not-so-perfect” users, the “perfect” users were significantly more likely to have reported at least one pregnancy during the trial [36]. This counter-intuitive result suggests that participants who consistently report 100% product adherence are, in fact, likely to be over-reporting adherence. Because the CONRAD CS trial had collected sexual and adherence data by partner type, further trajectory analysis based on gel use data with primary-partners-only was able to discern an additional cluster of low sustained gel users. Similar analysis from the MIRA trial, which had enrolled mainly monogamous married women, also identified four informative clusters, one consisting of 31% of women classified as low sustained diaphragm users [26, 37].

Recommendations

This paper drew from multiple sources to synthesize what has been learned, at the bench and in the clinic, about potential successes and limitations in achieving and measuring microbicide adherence in clinical trials. We conclude with a set of six recommendations that we hope may stimulate further improvements in both the measurement and optimization of adherence for microbicide development, and the field of biomedical HIV prevention interventions more broadly.

Establish Clarity of Purpose a priori

Not all microbicide trials are measuring the same things, particularly with respect to adherence, and consensus continues to evolve around what should be measured, why, how, and how often. Despite this variability, in all cases it is essential to be clear a priori about the purpose, or purposes, for which adherence-related data are to be collected.

Incorporate Adherence Measurement into Trial Design and Analysis

Although the ITT analysis is, and will likely remain, the gold standard for evaluating effectiveness in microbicide trials, adherence-adjusted analyses offer the potential for more powerful tests and richer understanding of trial data and their implications. Recent advances in causal inference may contribute greatly here, but more work is required to fulfill this potential, notably by identifying pre-randomization factors associated with good trial-related product adherence and more accurate assessments of adherence in relation to exposure to HIV. Like the a priori conceptualization of the role of adherence analyses in general, the possibilities for adherence-adjusted analysis must also be considered early in planning any trial or trial design alternative.

Develop, Test, and Validate New Measurement Approaches

More work is needed to develop, test and validate new approaches to measuring adherence. One priority area is the development of robust, validated biomarkers of microbicide safety and adherence, requiring further work to identify microbicide-induced changes in mucosa; assess the impact of reproductive hormones, microflora, and seminal plasma on microbicide-mucosa interactions; and determine the effects of repeated microbicide exposure on mucosa. As the microbicide field focuses more intensely on ART-based and coitally-independent products such as vaginal rings, prospects for electronic or biochemical

methods for monitoring frequency of product use and dosage seem more realistic and should be pursued.

A second priority area relates to the improvement of indirect and self-reported approaches to adherence measurement. This includes developing guidelines for tailoring adherence-related questions (e.g., time references, formats and data collection modes) to study populations; pursuing efforts to identify and validate a composite measure of adherence that would permit adherence comparisons across trials; and improving strategies to identify and resolve inconsistencies in reported sexual and adherence behavior through triangulation techniques.

Optimize Adherence in Trial Settings

Optimizing adherence within the trial setting is of crucial importance. Some strategies can be planned ahead. For example, a run-in period can be designed, perhaps utilizing the period between screening and enrollment, to collect more baseline information on potential trial populations with respect to the variables that are likely to be of primary importance to study success. Counseling messages and approaches should be shaped, tested, and modified to be understood as a package prior to study implementation. Explicit systems may be developed to support participants' on-going adherence including (1) motivational enhancement for those pre-identified as having adherence deficits, (2) provision of personal diaries; (3) use of adherence "buddies" or partners, and (4) individualized feedback on observed adherence patterns. However, given the limited availability of research to evaluate current approaches to optimizing adherence, future trials should also incorporate sufficient flexibility to monitor and adapt messages and/or strategies as needed.

Plan for Cross-trial Data Collection and Data Sharing

All trials cannot collect data on all aspects and contexts of adherence, but they may no longer need to do so. There is now a body of social science literature that could be compiled to serve all trials. It provides a deeper, more detailed understanding about the categories of sexual partners with whom topical gel use is or is not acceptable; the constellation of reasons why use varies temporally; and specific timing, dosage and insertion practices that might affect test product safety and effectiveness. Further, with respect to data sharing, when trials determine at the outset the purpose(s) for which they will use adherence data, it will be easier to develop strategies for sharing that information with other trials, planned or under way, thereby achieving economies of scale and helping accumulate a body of common knowledge for various applications of importance.

Articulate Guidelines for Reporting and Analyzing Adherence

Guidelines are needed on what to report and how to analyze adherence measures. Some recommendations include: conducting a thorough analysis of adherence patterns over time before deeming time-averaged summary measures as acceptable; carefully assessing potential biases of per-protocol analysis and performing randomization-based adherence-adjusted analyses when possible; and finally retaining—rather than censoring, participants who go off treatment or become pregnant in order to follow up their behavior and outcome.

As emphasized in a recent IOM report, future trials should consider partially-blinded factorial designs to evaluate the utility of adherence interventions to inform planning of future studies and forestall investments of time, effort, and funding of approaches that are of doubtful yield [10].

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. FHI. Phase 3 Trial in Nigeria Evaluating the Effectiveness of SAVVY Gel in Preventing HIV Infection in Women Will Close. 2006. <http://www.fhi.org/en/aboutfhi/media/releases/phase3savvy082806.htm>. Accessed 11 November 2009.
2. Skoler-Karpoft S, Ramjee G, Ahmed K, Altini L, Plagianos MG, Friedland B, et al. Efficacy of Carraguard for prevention of HIV infection in women in South Africa: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2008;372(9654):1977–87.
3. Stover K. Anti-HIV gel shows promise in large-scale study in women. NIH news (serial on the Internet). 2009. http://www3.niaid.nih.gov/news/newsreleases/2009/HPTN_035_gel.htm. Accessed 9 Feb 2009.
4. Van Damme L, Ramjee G, Alary M, Vuylsteke B, Chandeying V, Rees H, et al. Effectiveness of COL-1492, a nonoxynol-9 vaginal gel, on HIV-1 transmission in female sex workers: a randomised controlled trial. *Lancet*. 2002;360:971–7.
5. Van Damme L, Govinden R, Mirembe FM, Guédou F, Solomon S, Becker ML, et al. Lack of effectiveness of cellulose sulfate gel for the prevention of vaginal HIV transmission. *N Engl J Med*. 2008;359(5):463–72.
6. MDP. MDP 301 Completion of follow up web announcement v1.5. 2009. <http://www.mdp.mrc.ac.uk/>. Accessed 28 Sept 2009.
7. Mihalko SL, Brenes GA, Farmer DF, Katula JA, Balkrishnan R, Bowen DJ. Challenges and innovations in enhancing adherence. *Control Clin Trials*. 2004;25(5):447–57.
8. Ockene IS, Hayman LL, Pasternak RC, Schron E, Dunbar-Jacob J. Task force #4-adherence issues and behavior changes: achieving a long-term solution. 33rd Bethesda Conference. *J Am Coll Cardiol*. 2002;40(4):630–40.
9. Robiner WN. Enhancing adherence in clinical research. *Contemp Clin Trials*. 2005;26(1):59–77.
10. Lagakos S, Gable A. Methodological challenges in biomedical HIV prevention trials. Washington, DC: The National Academics Press; 2008.
11. Padian NS, van der Straten A, Ramjee G, Chipato T, de Bruyn G, Blanchard K, et al. Diaphragm and lubricant gel for prevention of HIV acquisition in southern African women: a randomised controlled trial. *Lancet*. 2007;370(9583):251–61.
12. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J Natl Cancer Inst*. 2007;99(13):1036–43.
13. Vandebosch A, Goetghebeur E, Van Damme L. Structural accelerated failure time models for the effects of observed exposures on repeated events in a clinical trial. *Stat Med*. 2005;24(7):1029–46.
14. Niedbala RS, Mauck C, Harrison P, Doncel GF. Biomarker discovery: validation and decision-making in product development. *Sex Transm Dis*. 2009;36(3 Suppl):S76–80.
15. Keller MJ, Herold BC. Understanding basic mechanisms and optimizing assays to evaluate the efficacy of vaginal microbicides. *Sex Transm Dis*. 2009;36(3 Suppl):S92–5.
16. Cummins JE, Jr., Doncel GF. Biomarkers of cervicovaginal inflammation for the assessment of microbicide safety. *Sex Transm Dis*. 2009;36(3 Suppl):S84–91.
17. Mauck CK, Straten A. Using objective markers to assess participant behavior in HIV prevention trials of vaginal microbicides. *J Acquir Immune Defic Syndr*. 2008;49(1):64–9.
18. Mauck CK. Biomarkers of semen exposure. *Sex Transm Dis*. 2009;36(3 Suppl):S81–3.
19. Simoni J, Kurth A, Pearson C, Pantalone D, Merrill J, Frick P. Self-report measures of antiretroviral therapy adherence: a review with recommendations for HIV research and clinical management. *AIDS Behavior*. 2006;10(3):227–45.
20. Liu H, Golin CE, Miller LG, Hays RD, Beck CK, Sanandaji S, et al. A comparison study of multiple measures of adherence to HIV protease inhibitors. *Ann Intern Med*. 2001;134(10):968–77.
21. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. *Psychol Bull*. 1990;108(3):339–62.
22. Friedland B. Adherence in the Carraguard Phase 3 trial. Adherence and its Measurement in Microbicide Clinical Trials; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
23. Noar SM, Cole C, Carlyle K. Condom use measurement in 56 studies of sexual risk behavior: review and recommendations. *Arch Sex Behav*. 2006;35(3):327–45.
24. van der Straten A. Methodological challenges and measurement issues in MIRA. Adherence and its Measurement in Microbicide Clinical Trials; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
25. Tolley EE, Tsui S, Kohli R, Weaver M. HPTN 059 Sub-study: explaining differential condom use among clinical and non-clinical trial cohorts. Adherence and its Measurement in Microbicide Clinical Trials; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
26. Van der Straten A. Produce use and adherence. Adherence and its Measurement in Microbicide Clinical Trials; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
27. Mensch B, Rankin J. Assessing reporting of adherence and sexual behavior in a simulated microbicide trial: preliminary results from an interview mode experiment in the Phase 3 Carraguard sites, South Africa. Adherence and its Measurement in Microbicide Clinical Trials; December 18, 2007. Washington, DC: Alliance for Microbicide Development; 2007.

28. Gorbach P, Mensch B, Coly A, Mierwa S, Cao S, Sites MM. Pictorially-based AcASI data collection in a microbicide clinical trial. *Adherence and its Measurement in Microbicide Clinical Trials*; December 18–19, 2009. Washington, DC: Alliance for Microbicide Development; 2007.
29. Carballo-Diéguez A, Ventuneac A. Interactive voice response (IVR) data collection mode & related technologies. *Adherence and its Measurement in Microbicide Clinical Trials*; December 18–19, 2009. Washington, DC: Alliance for Microbicide Development; 2007.
30. McCormack S, Pool R. Measuring adherence in MDP 301. *Adherence and its Measurement in Microbicide Clinical Trials*; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
31. Taylor D, Richardson B. Role of adherence measurement in determining effectiveness. *Adherence and its Measurement in Microbicide Clinical Trials*; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
32. Deese J, Van Damme L. FHI Truvada study 10015. *Adherence and its Measurement in Microbicide Clinical Trials*; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
33. Mansoor LA. Overview of the Adherence Support Program in CAPRISA 004 Phase IIb Tenofovir Gel Trial. *Adherence and its Measurement in Microbicide Clinical Trials*; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
34. Coletti AS. Phase II/IIb Safety and effectiveness study of the vaginal microbicides BufferGel and 0.5% PRO 2000/5 Gel (P) for the prevention of HIV infection in women. *Adherence and its Measurement in Microbicide Clinical Trials*; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
35. Vrijens B, Goetghebeur E. Comparing compliance patterns between randomized treatments. *Control Clin Trials*. 1997;18: 187–203.
36. Weaver M, Tolley EE, Tsui S. Patterns of adherence in two Phase III microbicide trials. *Adherence and its Measurement in Microbicide Clinical Trials*; December 18–19, 2007. Washington, DC: Alliance for Microbicide Development; 2007.
37. van der Straten A, Shiboski S, Montgomery ET, Moore J, De Bruyn G, Ramjee G, et al. Patterns and predictors of adherence to diaphragm use in a Phase III trial in sub-Saharan Africa: a trajectory analysis. *J Acquir Immune Defic Syndr*. 2009;50(4): 419–26.